

UC Riverside

UC Riverside Previously Published Works

Title

Connecting species' geographical distributions to environmental variables: range maps versus observed points of occurrence

Permalink

<https://escholarship.org/uc/item/5j41c8sv>

Journal

Ecography, 43(6)

ISSN

0906-7590

Authors

Rotenberry, JT
Balasubramaniam, P

Publication Date

2020-06-01

DOI

10.1111/ecog.04871

Peer reviewed

ECOGRAPHY

Research

Connecting species' geographical distributions to environmental variables: range maps versus observed points of occurrence

John T. Rotenberry and Priya Balasubramaniam

J. T. Rotenberry (<https://orcid.org/0000-0002-0864-1676>) ✉ (jrotenbe@umn.edu), Dept of Ecology, Evolution, and Behavior, Univ. of Minnesota, Saint Paul, MN, USA. – P. Balasubramaniam, Reno, NV, USA.

Ecography

43: 1–17, 2020

doi: 10.1111/ecog.04871

Subject Editor: Jorge Soberon

Editor-in-Chief: Hanna Tuomisto

Accepted 13 February 2020



Connecting the geographical occurrence of a species with underlying environmental variables is fundamental for many analyses of life history evolution and for modeling species distributions for both basic and practical ends. However, raw distributional information comes principally in two forms: points of occurrence (specific geographical coordinates where a species has been observed), and expert-prepared range maps. Each form has potential short-comings: range maps tend to overestimate the true occurrence of a species, whereas occurrence points (because of their frequent non-random spatial distribution) tend to underestimate it. Whereas previous comparisons of the two forms have focused on how they may differ when estimating species richness, less attention has been paid to the extent to which the two forms actually differ in their representation of a species' environmental associations. We assess such differences using the globally distributed avian order Galliformes (294 species). For each species we overlaid range maps obtained from IUCN and point-of-occurrence data obtained from GBIF on global maps of four climate variables and elevation. Over all species, the median difference in distribution centroids was 234 km, and median values of all five environmental variables were highly correlated, although there were a few species outliers for each variable. We also acquired species' elevational distribution mid-points (mid-point between minimum and maximum elevational extent) from the literature; median elevations from point occurrences and ranges were consistently lower (median –420 m) than mid-points. We concluded that in most cases occurrence points were likely to produce better estimates of underlying environmental variables than range maps, although differences were often slight. We also concluded that elevational range mid-points were biased high, and that elevation distributions based on either points or range maps provided better estimates.

Keywords: elevation, environmental variables, Galliformes, points of occurrence, range maps, species geographical distributions

Introduction

Animal life histories are shaped by the environmental attributes associated with a species' geographical distribution and by any phylogenetic constraints, the latter associated with the species' evolutionary history (Roff 1992). As information concerning



www.ecography.org

© 2020 The Authors. Ecography published by John Wiley & Sons Ltd on behalf of Nordic Society Oikos
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

life history, distribution and phylogeny becomes increasingly available for a variety of taxa, it is now possible to imagine a global analysis of life history patterns encompassing the majority of species in many major clades (Jetz et al. 2008, for clutch size variation across over 5000 species of birds). We are particularly interested in environmental influences on variation in avian life histories that vary along latitudinal and elevational gradients (Lack 1947, Boyle et al. 2016), motivated by our previous finding that elevation and latitude interact to influence several measures of reproductive effort in galliform birds (Balasubramaniam and Rotenberry 2016). Whereas most analyses treat elevation (measured in meters) and latitude (measured in degrees north or south of the equator) as proxies for variation in a suite of environmental variables (Jetz et al. 2008 is an exception in this regard), better understanding of the role of extrinsic drivers of life history evolution will come from replacing meters and degrees with actual values of environmental variables (e.g. annual precipitation or net primary productivity, annual temperature and its seasonality, hours of daylight during the breeding season, etc.) associated with species' distributions into our analyses. This will allow us to test alternative theories of the drivers of life history evolution along such gradients more directly.

Ascertaining environmental attributes associated with a species' distribution is also relevant for many other ecological applications, most particularly in determining and/or modeling species-habitat relationships (Franklin 2010). Many studies have a strong ecological focus, examining causal drivers of species' distributions and inferring niche relationships (Elith and Leathwick 2009), and some address evolutionary questions such as mechanisms of speciation (Graham et al. 2004). Frequent practical applications of these relationships include species and habitat conservation planning, reserve design, habitat management and predicting species' responses to environmental change (Gusian et al. 2013, Franklin et al. 2014).

Regardless of the conceptual nature of the research question, the critical first step in such analyses is determining the geographical distribution of a species, which may then be connected to spatially explicit maps of environmental variables. Bird species distributions, as well as those of many other taxa, are generally available in two forms: range maps, and collections of points (at a much finer scale of spatial resolution than a range map) where a species has been observed. Range maps, also called extent-of-occurrence maps, are created by drawing one or more polygons that encompass all known or inferred locations of a species' occurrence, and are usually prepared by experts based on their own knowledge of the species' distribution (IUCN 2016). Depending on the scale of resolution, edges of a range polygon may appear either smooth or irregular, and the interior may be either completely filled or have holes indicating known or expected gaps in the species' distribution. Point locations may be collected from a variety of sources (e.g. surveys, atlases, citizen-science reporting such as eBird (Sullivan et al. 2009, 2014), museum records) and include records with varying degrees of spatial precision.

Conspicuous, widely-distributed species may have millions of records in a global database (GBIF.org 2018).

Both forms of data have potential shortcomings in describing the details of a species' geographical distribution. Range maps often overestimate the true occurrence of a species; that is to say, the species is not found everywhere within the boundaries of the range polygon(s). This 'range porosity' (Hurlburt and White 2005) may be due to discontinuities in suitable habitat or simply reflect a naturally patchy distribution of individuals. Moreover, range boundaries are usually drawn as relatively smooth and simple shapes, which may fail to capture the irregular distribution of species at its range limits (Fortin et al. 2005). At the same time, range maps may also leave out some areas of occupancy associated with peripheral or disjunct populations. On the other hand, occurrence points are usually non-randomly distributed across a species' potential distribution, often spatially clumped (e.g. birding 'hot spots,' or along roads or other features that facilitate ease of access), and the geographical coordinates themselves may be imprecise. Moreover, geographically precise point data may be scant or missing for many species, particularly those of limited ranges that occur in relatively difficult-to-access areas.

Elevational data pose a somewhat different problem. In association with a range map, a species' elevational distribution is almost invariably described as a span between lower and higher limits within which the species may occur, and usually reported at a resolution no finer than the nearest 100 m, and often more coarse (del Hoyo et al. 2017, Quintero and Jetz 2018). Moreover, reported limits for a species may vary among different mountain ranges within a species' distributional range (Quintero and Jetz 2018). Absent any modifying information, if a single number is used to describe a species' elevation, it is usually the mid-point of the range limits (Boyce et al. 2015, Balasubramaniam and Rotenberry 2016). Occurrence points, on the other hand, may be coupled with more precise elevational data associated with each point, with the degree of precision depending on the scale of resolution of the point itself and of the underlying elevational data.

In application, distributions based on range maps may be modified by deleting areas that represent habitat types in which the species is not known to occur or areas outside the elevational limits of the species (Scott et al. 1993, Jenkins et al. 2013). Occurrence points considerably disjunct from a species known range or outside habitat and elevational limits may be deemed unreliable or irrelevant and deleted (see below). Such modifications assume, of course, that these ranges, relevant habitat types and elevational limits are correctly identified for the species to which they are applied (Peterson et al. 2016).

These differences in attributes of the two principal sources of distributional data have prompted a variety of studies to assess how each influences analyses based on one or the other. A number of studies have focused on geographical variation in species richness, often in the context of conservation planning. A consistent result is that, compared to surveys or other

collections of point-based data, range maps over-estimate species' occupancy, and hence overlaying range maps over-estimates species richness at any particular place compared to on-the-ground surveys of the same area (Hurlburt and White 2005, Graham and Hijmans 2006, Hurlburt and Jetz 2007, McPherson and Jetz 2007, Jetz et al. 2008). This effect is scale-dependent, with the discrepancy between range-based and point-based richness estimates increasing with finer spatial resolution. Thus, for any particular point a range map may create a 'false-positive' of a species' occurrence. Other studies have examined the 'completeness' of species data from points, again comparing surveys or site-specific species lists with the presence of species in an area determined by reported occupancy points. These, too, find discrepancies, principally that not all species known to be present in an area are necessarily represented in available point of occurrence data (Yesson et al. 2007, Jacobs and Zipf 2017, Qian et al. 2018). Again, this effect is scale-dependent; the discrepancy decreases as the size of the area surveyed increases. Such an effect can generate a 'false-negative' of a species' occurrence, to the extent a point where a species has not been detected is assumed for the purposes of any analysis to be unoccupied.

What remains less investigated is the extent to which range maps and point-based data actually differ in their representation of a species' environmental associations (Alhajeri and Fourcade 2019). For example, if a range map polygon contains a considerable proportion of otherwise unsuitable environmental conditions for a species, then we expect to see a discrepancy between mean or median values of the relevant variables calculated over the entire range versus those calculated just from known occupied points within the range. We might expect, then, that this discrepancy increases with greater implicit (as opposed to explicit) porosity of the range map. Another source of discrepancy may be the distributional bias that can occur in point data, as points may not be randomly distributed throughout the actual area of occupancy of a species (Boakes et al. 2010, Beck et al. 2014, Fourcade et al. 2014, Fourcade 2016, Meyer et al. 2016). Regardless, to the extent that environmental associations do differ between the two, we can expect to see differences in conclusions we might draw about life history patterns, or for that matter in any other macroecological analysis that relies on these associations (e.g. species distribution modeling; Beck et al. 2014, Fourcade 2016, Alhajeri and Fourcade 2019).

With the foregoing in mind, we pose two questions. First, do estimates of species' geographical distributions (longitude, latitude) and associated environmental (i.e. climate) variable values differ whether based on range maps versus points? To the extent that there is a difference in the centroid or dispersion of a species' geographical distribution between the two sources, we might expect any differences in environmental variables across species to scale with the distance between centroids. To the extent that range maps may include inappropriate as well as appropriate environmental conditions for a species, we might further expect that the variances of environmental variables calculated for species' range maps

should be greater than those calculated based on species' observed occurrences. Alternatively, if we include occupied points that lie outside range map boundaries (see below) we might observe a greater variation associated with point estimates than for range maps. Second, do estimates of species' elevational distributions differ whether based on range maps versus points versus mid-points of reported spans (minimum to maximum)? Again, we expect that differences between range maps and points might scale with distances between centers of their geographical distributions, and that elevations based on maps may have greater variances than those based on occurrences. Although it was not clear a priori what we might predict regarding maps or points versus mid-points of elevational spans, we subsequently observe a substantial difference between the first two and the latter.

We address these questions using a globally distributed taxon, the avian order Galliformes, consisting of the pheasants, quails, guineafowl, guans and megapodes. The order includes 294 species in five families (IOC World Bird List V.7.3, IOC hereafter; Gill and Donsker 2017). Galliforms span a large size range (20–11 200 g), occur on all continents (except Antarctica) and in virtually all biome types, have a wide latitudinal extent (from 43°S to more than 80°N), and may be found from below sea level to over 6000 m (summarized in Handbook of birds of the world online, HBW hereafter; del Hoyo et al. 2017). Our estimates of a species' area occupied (in our case, the number of 10' × 10' map cells) based on extent-of-occupancy range maps ranged from 1 to over 163 000 (approximately 25% of ice-free terrestrial land), and from 1 to 17 000 based on point distributions (see below). Thus galliforms should be reasonably representative of the extent of variation we might expect to see in terrestrial birds.

Material and methods

Environmental data

We used climate data from WorldClim2 (Fick and Hijmans 2017) at a resolution of 10' (0.16667°), which yielded a global terrestrial raster layer of 808 053 cells (583 798 cells excluding Antarctica). Each cell is approximately 18.4 km on a side at the equator, with an area of about 340 km². We used this base layer throughout our analyses. Elevation data were taken from GTOPO30 (Global Digital Elevation Model 2004) at an original resolution of 30 arc-seconds resampled to 10' centered on the base layer cells. Each environmental variable was then imported into ModestR (<www.ipez.es/ModestR/>), a convenient tool for managing species distributional data in either point locations or range map format, and environmental data in raster format (Garcia-Rosello et al. 2013, 2014).

Species distribution data

The source of our primary point-of-occurrence data was the Global Biodiversity Information Facility (GBIF.org 2018;

<www.gbif.org/>; GBIF hereafter), which amalgamates occurrence data from thousands of sources, including data from eBird (Sullivan et al. 2009, 2014). One or more species of interest is specified, and then points for each are downloaded. Species data for these analyses were downloaded 16 January–4 May 2018, and links to the raw data we downloaded are provided in Supplemental material Appendix 1 Table A1. Any differences between GBIF species taxonomy and IOC were resolved using IOC taxonomy, the principal differences being recent changes to genus names that had yet to be incorporated into all data sourced from GBIF. Points described as ‘invalid record’ or ‘invalid number of species’ were deleted. The latter sometimes occurs when all locations (i.e. survey blocks) in an atlas compendium are uploaded into GBIF, including locations where the species of interest is absent (the number of the target species is zero, an invalid number for presence data). Likewise, points with missing longitude and latitude were deleted. Data were sorted by longitude and latitude, and then all but one of any points with the same coordinates were deleted so that there were no duplicate coordinates.

Species’ point data were then imported into ModestR. First, any point that fell into water (generally ocean) was deleted. Then, a range map from BirdLife International/NatureServe (BirdLife International and NatureServe 2015, BirdLife hereafter; <<http://datazone.birdlife.org/home>>) was uploaded (see below). Point locations were overlaid on the range map and any points > 500 km from the edge of the breeding range were deleted. We retained points within 500 km from range edges to account for potential peripheral or disjunct populations omitted from range boundary determination. Finally, points representing known introductions of a species outside its native range were deleted. Areas of introduction were derived from information presented in HBW (del Hoyo et al. 2017). In a similar analysis for mammals, Alhajeri and Fourcade (2019) created two data sets from GBIF, one containing all points after cleaning regardless of proximity to the range edge or status as introductions, the other retaining only those species’ points that fell within range polygons.

Cleaned points were exported in a spreadsheet, then rasterized onto our global terrestrial raster layer at a resolution of 10’ using library Raster in R (R Core Team, Hijmans 2015). Each cell in the raster layer was then denoted as occupied by a species if it contained one or more species’ points, regardless of the number of points. Otherwise a cell was unoccupied. The rasterized data were then reimported into ModestR.

Shapefiles of species range maps from BirdLife were imported directly into ModestR at a resolution of 1’. Any differences between BirdLife taxonomy and IOC were resolved using IOC taxonomy; as with data imported from GBIF, most differences involved genus reassignments. Unlike for points, rasterization of range maps took place within ModestR. Using the same global 10’ base map, a cell was denoted as occupied if at least 25% of it was included within a species’ range boundary (i.e. at least 25 1’-cells within the 10’ × 10’ cell fell within the species’ range). Because maps

from shapefiles cannot be edited in ModestR, to clean the map further we exported the rasterized data then reimported them as a point file. As with processing GBIF data, we removed any points that were in water, and those from areas of introduction (range maps varied in whether they included these areas). If the range map included non-breeding distribution, we removed those points as well. Note that BirdLife is the authority for IUCN Red List maps for birds (IUCN 2016), a primary source for other vertebrate and invertebrate distributions.

We used two additional data sets containing information on the elevational distribution of each species. Quintero and Jetz (Q&J hereafter; 2018) summarize the elevational ranges (minimum and maximum elevation) of 9993 species based on 318 published sources (about 70% from HBW), and from that we calculated the elevational mid-point for 263 galliforms. We also downloaded elevational range limits from BirdLife for 281 galliforms. However, a number of these were missing either upper or lower limits, which we filled in using species accounts in HBW. For those still missing an explicit lower limit, we used zero in those cases where it was evident that the species range extended to or near the coast (sea-level) based on the species account in HBW. This yielded 242 species for which we could calculate an elevational mid-point.

Analysis

After overlaying a species’ distribution on raster maps of environmental variables, we output one file containing values for longitude, latitude, elevation and four representative climate variables for each occupied cell for each species based on its point locations, and a similar file for occupied cells based on range maps. Due to spatial autocorrelation (and the fact that some variables are derived from others) many of the 19 WorldClim2 climate variables are strongly covariant at the scale of our analysis; for example, annual precipitation was highly correlated with amount of precipitation in the wettest quarter of the year ($r=0.94$, $n=808\,053$ cells). Rather than examine all 19 variables, we selected four representative of the magnitude and variability of temperature and precipitation: 1) annual mean temperature, 2) total annual precipitation, 3) mean diurnal temperature range (mean of monthly maximum temperature minus minimum temperature) and 4) precipitation seasonality (coefficient of variation of monthly precipitation). These four variables are among the most frequently used for species distribution modeling (Gardner et al. 2019). See Supplemental material Appendix 1 Table A2 for more detailed descriptions of these variables and their intercorrelations.

We then calculated median values of each variable for each species based on its point locations and on its range map. We chose medians rather than means as representative of central tendency in these data because the frequency distributions of most of the variables we examined were decidedly non-normal (and in the event a variable is normally distributed the mean and median converge). Moreover, the median is a ‘resistant’ statistic (Zar 1984, p. 22) that is much less

influenced by outlying or extreme values than is the mean. Similarly, we used the interquartile range (the span between the 25th and 75th percentiles) and interdecile range (the span between the 10th and 90th percentile) as robust measures of the variation of each variable, the former less sensitive to outliers or extreme values than the latter. We then calculated the correlation between point-based and range-based medians and variances for each variable across all species to estimate the strength of the relationship between the two estimates. To supplement our interpretation of these relationships, we regressed point-based values on range-based ones (as a matter of convenience, not to imply any functional dependency between the two); although two variables might be perfectly correlated, a substantial difference from 1 in the slope of the relationship suggests that values of one variable are not necessarily equal to values of the other variable across the full range of variable values. We also calculated the median absolute difference between point-based and range-based values for a variable across all species, and correlated those differences with the approximate great-circle distance between point-based and range based geographical centroids (centroid defined as the median longitude, median latitude for a species). We calculated the great-circle distance (the distance between two points on the surface of a sphere) between range-based and point-based geographical centroids using the haversine formula (van Brummelen 2013).

We performed similar correlations and regressions comparing elevation estimates from Quintero and Jetz (2018) and Birdlife/NatureServe to each other and to those elevations we generated based on range maps and occurrence points.

Because species distributed across small spatial extents likely experience a smaller range of environmental conditions, it is possible that their correlations between range-based and point-based medians might be higher (Alhajeri and Fourcade 2019); thus we repeated these correlations within each of three groups of species based on the number of $10' \times 10'$ cells included within their range maps. As there were no obvious breaks in the frequency distribution of range sizes, particularly at sizes $< 10\,000$ cells, we simply classified them as small (1–1000 cells, $n = 132$ species), medium (1000–10 000 cells, $n = 117$ species) and large ($> 10\,000$ cells, $n = 40$ species). Identity of species in each category can be found in data deposited in Dryad. Because of the absence of clear-cut range size categories we also calculated partial correlations of range-based versus points-based variables, partialling out covariation with range size.

We conducted all correlations and regressions in SAS 9.3 (SAS Inst. 2013).

Results

Longitude, latitude, climate variables

Median geographical location of a species' distribution varied little between latitude and longitude determined from range

maps versus reported point locations (Table 1, Fig. 1); indeed, over all species the correlation between the two data sources was 0.99 for longitude and 0.98 for latitude, with regression coefficients of 0.99 and 1.00 respectively. There was a very slight tendency for ranges to have a southern bias compared to medians based on points (median point latitude minus range latitude = -0.17°) although the median difference in longitude = 0.00° . The median distance between geographical centroids based on points versus ranges was 234 km, ranging from zero to nearly 5400 km (Fig. 2).

The general agreement in geographical distributions (at least as indicated by similarity of median longitude and latitude, and the distance between centroids) yielded comparably high correlations in estimated annual temperature and annual precipitation ($r = 0.97$ and 0.96 , respectively), albeit with regression slopes (0.88 and 0.93 , respectively) slightly lower than those for longitude and latitude (Fig. 3A–B). Likewise, diurnal temperature range and precipitation seasonality were highly correlated across the two types of distributions (both $r = 0.94$, slopes = 0.93 and 0.97 , respectively; Fig. 3C–D).

Differences in environmental medians based on points versus those based on ranges (i.e. the distance from a point to the 1:1 line in Fig. 3) were not well explained by the geographical distance between species' centroids based on the two different data sources. Regressing absolute values of the difference between median annual temperature from points and from ranges on distance between centroids yielded $r^2 = 0.11$ (slope = 0.08°C per 100 km). Similar regressions for the other variables yielded $r^2 = 0.02$ for annual precipitation (slope = $3.5\text{ mm}/100\text{ km}$), 0.12 for diurnal temperature range (slope = $0.03^\circ\text{C}/100\text{ km}$), and 0.18 for precipitation seasonality (slope = $0.06/100\text{ km}$).

Variances for these variables, as represented by interquartile and interdecile ranges, were also highly correlated between ranges and points, although somewhat less so than medians, and varied more across variables than did correlations of medians (Table 2). Correlations between range- and point-based interdeciles were slightly greater than those for interquartiles across all variables, and except for elevation the median absolute differences between points and ranges were higher for interdeciles than interquartiles. Interquartile and interdecile ranges were greater for range-based data than point-based data for all of these variables except elevation and mean annual temperature. For range-based data, correlations of interquartile and interdecile values for the seven variables ranged from 0.93 to 0.99 across species, and for point-based data they ranged from 0.83 to 0.97 (Table 2).

Range size had essentially no effect on correlations of medians between the two data sources (Table 3); there were scant differences in correlations across range size classes, and partial correlation coefficients were nearly identical to original correlations not accounting for range size variation. Range size had some effect on variance estimates, more so on interquartile ranges than interdecile ones (Table 3). Based on differences between partial correlations and original correlations longitude seemed most affected.

Table 1. Basic statistics, correlations and regressions comparing median environmental variable values generated from range-based versus points-based species' distributions. n = 291 species for points, 289 species for ranges, 286 for differences, correlations and regressions.

	Basic statistics					
	Range-based			Points-based		
	Median	Minimum	Maximum	Median	Minimum	Maximum
Number of cells	1354	1	163172	53	1	16997
Longitude (degrees)	39.75	-128.75	167.42	34.92	-122.92	167.08
Latitude (degrees)	9.58	-32.59	65.92	9.58	-33.25	64.92
Elevation (m)	521.2	19.9	4624	583.6	21.4	4350
Mean annual temperature (C)	23.0	-9	28	21.5	-4	28
Annual precipitation (mm)	1420	93	3596	1419	133	3707
Diurnal temperature range (C)	10.12	5.38	17.34	9.87	5.87	17.31
Precipitation seasonality (no units)	63.15	10.44	149.34	61.76	10.49	138.52
						4.05

	Regression (points = intercept + slope × range)	
	Correlation	Slope
		Intercept
Number of cells	0.47	201.9
Longitude (degrees)	0.99	-0.96
Latitude (degrees)	0.98	-0.87
Elevation (m)	0.95	106.2
Mean annual temperature (C)	0.97	2.06
Annual precipitation (mm)	0.96	128.9
Diurnal temperature range (C)	0.94	0.51
Precipitation seasonality (no units)	0.94	0.36

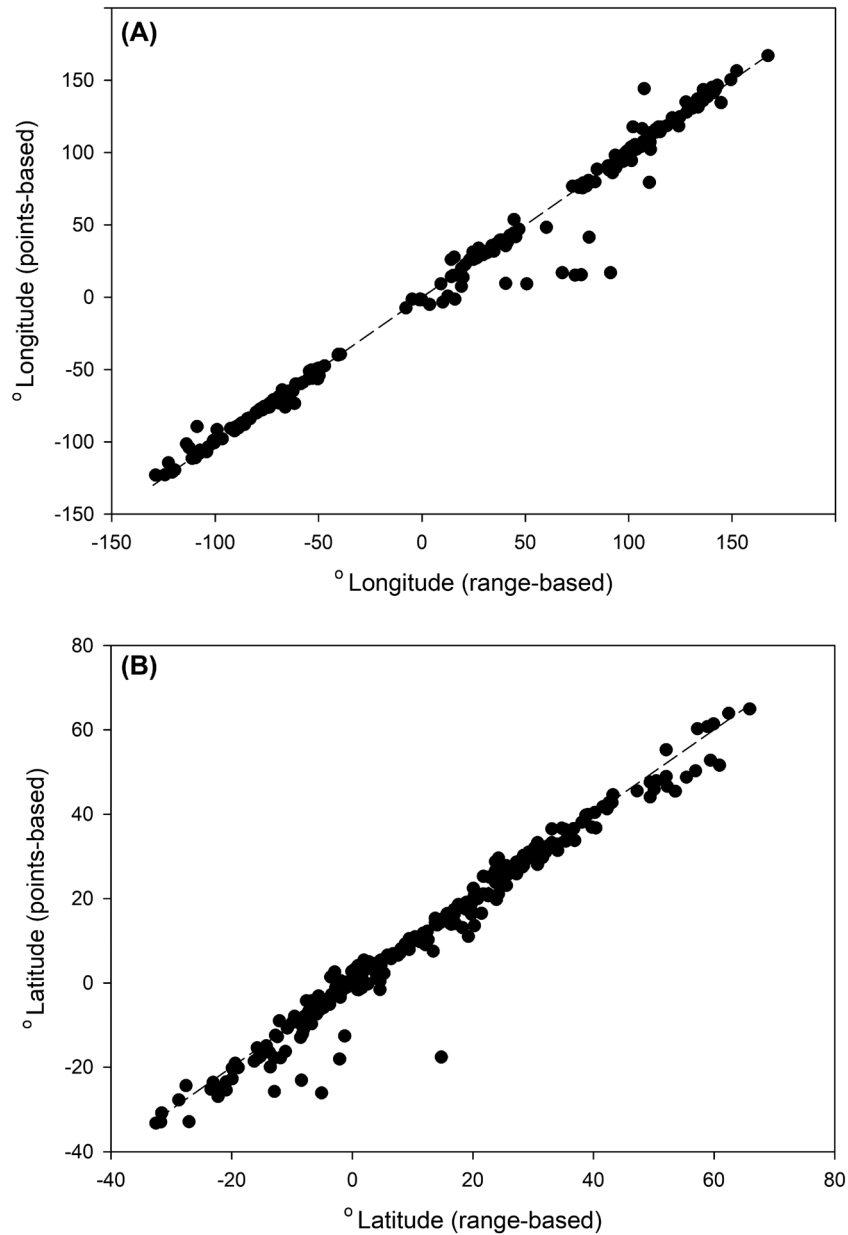


Figure 1. Median longitude (A) and latitude (B) derived from points versus medians derived from range maps. Dashed line represents 1:1 relationship. See Table 1 for correlation and regression coefficients.

Elevation

As with the other environmental variables, there was a strong relationship ($r=0.95$, slope=0.95) between median elevations derived from ranges versus those derived from points (Table 1, Fig. 4A). At the same time, some species showed considerable discrepancy between the two estimates (in a few cases as much as 1000 m), although overall the median absolute difference was 120.5 m. Greater differences were associated with species that had relatively few points; the median absolute difference for 44 species with < 10 points was 254 m, whereas it was 107 m for 242 species with ≥ 10 points. These differences in elevations were not explained by the distance between centroids; regression

of absolute difference on distance yielded $r^2=0.001$. Similarly, the variation in elevation was correlated across the two estimates ($r=0.71$ for interquartiles, $r=0.82$ for interdeciles), with range-based interquartile and interdecile ranges mostly greater than point-based ones (Table 2). Based on differences between partial correlations and original correlations across all range sizes, median elevation correlations were independent of range size, as were those of interquartiles and interdeciles (Table 3). Overall, the differences were independent of the distance between centroids ($r=0.01$ for interquartiles, $r=-0.01$ for interdeciles). Because of the similarity of range-based and point-based estimates, we elected to use just the point-based data in subsequent analyses of elevation.

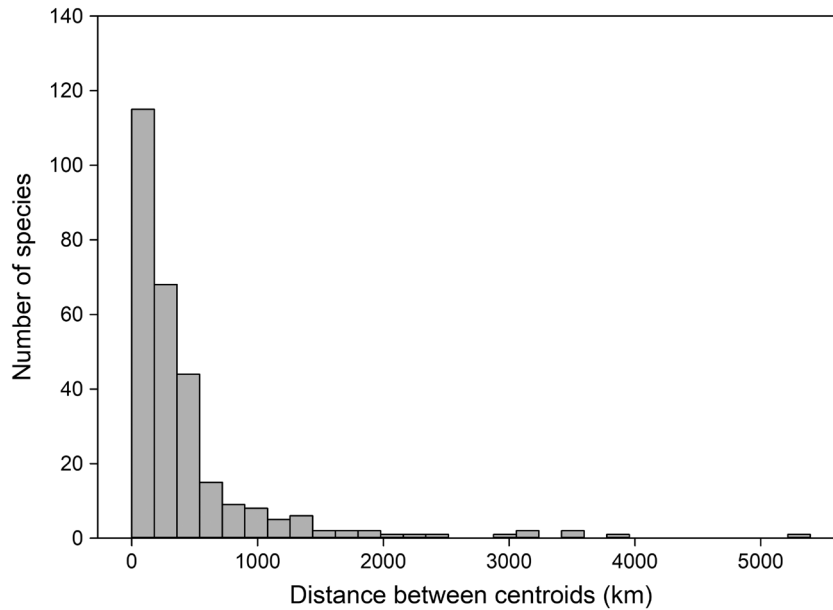


Figure 2. Distances between centroids of species geographical distributions determined from occurrence points and range maps.

Not surprisingly, since much of their data came from the same sources, there was a high correlation between mid-point elevations taken from Q&J and those from BirdLife ($r = 0.97$, 219 species) and a slope of Q&J regressed on BirdLife near 1 ($b = 0.98$). The median absolute difference between the two was 20.0 m (Q&J higher). Given this similarity, we chose to retain only Q&J estimates going forward because of the larger sample size (262 species versus 242 species).

Differences between elevation estimates based on mid-points between maxima and minima (Q&J) versus median of all occupied points were substantial, up to as much as 3400 m (Fig. 4B); the median absolute difference was 420 m (262 species), with estimates based on mid-points being higher. Although there was clearly a positive association between the two, the correlation ($r = 0.86$) was less than those observed in our analyses of the climatic variables. Of the 262 species, mid-points produced higher estimates of elevation than medians for 220. Differences between Q&J mid-points and medians of occupied points of ≥ 500 m occurred in 115 species, all but seven indicating higher elevations from Q&J. The slope of the regression of mid-point elevations on medians was 0.85 with an intercept of 570.3, which implies that the differences between mid-points and points-based estimates were higher at lower elevations (Fig. 4B).

Quintero and Jetz (2018), using N-mixture occupancy models applied to an extensive but locally intensive survey of birds in Switzerland (www.vogelwarte.ch/en/projects/monitoring/monitoring-common-breeding-birds), determined that field surveys consistently underestimated bird species richness at higher elevations. This implies that species' occurrences or presences may also be underestimated at higher elevations, potentially leading to a positive skew in the frequency distribution of a species' occurrence along an elevational gradient (i.e. relatively

fewer observations at higher elevations). Consistent with this expectation, of 263 species in our sample with at least 50 points, 233 had a skewness coefficient greater than zero (although some only marginally so).

Discussion

Although estimates of species' geographical distributions (longitude, latitude) and associated environmental (i.e. climate) variable values differed whether based on range maps versus points in ways we might have predicted, perhaps not as much as we might have expected. For the most part values of environmental variables derived from the two different sources in our analyses were highly correlated across species, which we might expect given the overlap in points and ranges due to the constraints we placed on including points for analysis (no further than 500 km from published range boundaries). These results mirror those of Alhajeri and Fourcade (2019) in a similar analysis based on data for 1191 species of rodents, who observed high correlations across all 19 WorldClim variables plus elevation in a comparison of GBIF points to IUCN range maps. Using all of a species' GBIF points regardless of range boundaries, Alhajeri and Fourcade (2019) observed a median distance between centroids of GBIF and IUCN data of 224 km, remarkably close to our observed 234 km. Moreover, in addition to using all of a species' GBIF points, Alhajeri and Fourcade (2019) repeated their analyses using only points falling within range map boundaries; environmental variable correlations were even higher for the constrained data, but only slightly. Although they assessed effects of geographic range sizes differently than we did, they did not observe differences in patterns of association across range sizes.

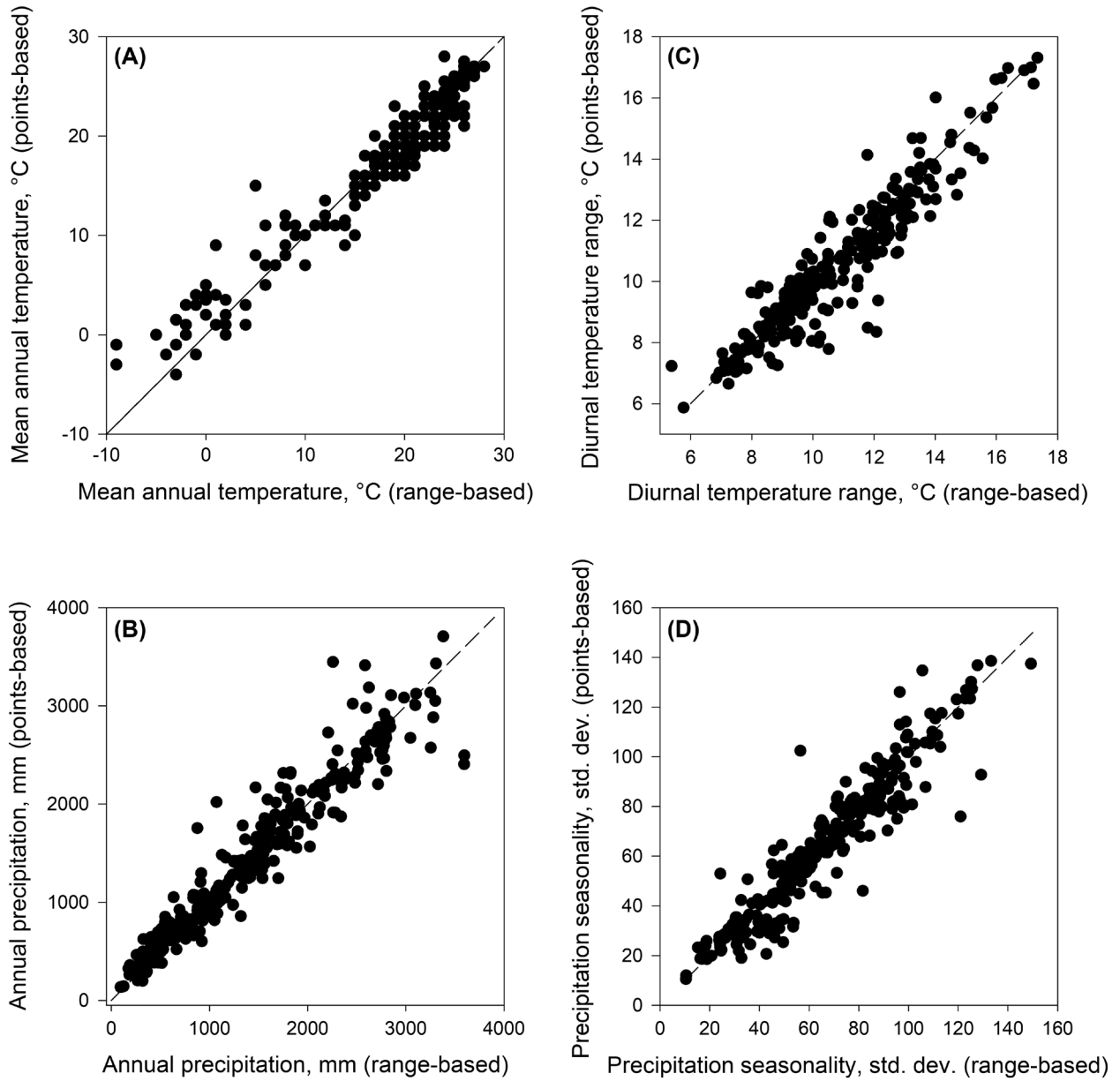


Figure 3. Medians of mean annual temperature (A), annual precipitation (B), diurnal temperature range (C) and precipitation seasonality (D) derived from points versus medians derived from range maps. Dashed line represents 1:1 relationship. See Table 1 for correlation and regression coefficients.

Although estimates of range-based versus point-based values for geographic and environmental variables were generally tightly associated, nevertheless some species differed considerably (or at least more than the large majority of species) between the two estimates. For example, there was a notable difference in longitude for several species (Fig. 1A) that showed an eastern bias in range-based estimates (or alternatively, a western bias in point-based estimates). This occurred in several species whose distributional ranges are reported to be broadly Palearctic (e.g. hazel grouse *Tetrastes bonasia*; black grouse *Lyrurus tetrix* (Fig. 5); western capercaillie *Tetrao urogallus*) or even circum-Arctic (willow grouse

Lagopus lagopus), but for whom there are very few records in north Asia east of 40°E. For willow grouse, only 11 of 2660 points are east of 34°E despite its ostensible circumpolar distribution. While it is difficult to attribute unequivocally a cause for the differences in longitude in the examples above, the scarcity of northeastern Asian observations for these grouse likely reflects the scarcity of any records at all for the region in question, rather than simply a regional scarcity of these species. Boakes et al. (2010) assessed the spatial distribution of Eurasian galliform locality records from a variety of sources, noting that whereas records obtained from museums were spread throughout the area of interest, northern Asia

Table 2. Basic statistics and correlations comparing median, interquartile and interdecile ranges of environmental variable values generated from range-based versus points-based species' distributions. n = 291 species for points, 289 species for ranges, 286 for differences and correlations.

	Interquartile ranges						
	Range-based			Point-based			Correlation
	Median	Minimum	Maximum	Median	Minimum	Maximum	
Longitude (degrees)	4.50	0	216.0	4.17	0	129.8	0.89
Latitude (degrees)	3.83	0	25.3	3.00	0	30.2	0.84
Elevation (m)	506.4	10.7	3162	593.9	1.8	2995	0.71
Mean annual temperature (C)	3.0	0	19	4.0	0	17	0.72
Annual precipitation (mm)	472	24.5	3014	460	11	2220	0.66
Diurnal temperature range (C)	1.67	0.17	4.26	1.52	0.04	4.75	0.75
Precipitation seasonality (no units)	20.35	0.02	86.14	18.37	0.02	71.75	0.73
	Interdecile ranges						
	Range-based			Point-based			Correlation
	Median	Minimum	Maximum	Median	Minimum	Maximum	
Longitude (degrees)	8.00	0	279.3	7.00	0	176.8	0.92
Latitude (degrees)	6.50	0	40.0	6.00	0	87.3	0.85
Elevation (m)	967.3	0	4427.0	1141.1	0	3703.2	0.82
Mean annual temperature (C)	6.0	0	25	7.0	0	21	0.80
Annual precipitation (mm)	893	0	5176.0	892	0	4048	0.84
Diurnal temperature range (C)	3.12	0	8.57	2.85	0	7.81	0.85
Precipitation seasonality (no units)	37.35	0	103.95	35.99	0	101.92	0.82

Table 3. Correlations comparing median, interquartile range and interdecile range of environmental variables generated from range-based versus points-based species' distributions, partitioned by range size (number of cells covered by IUCN range maps), and with range size partialled out. Range size groups are small (1–1000 cells), medium (1000–10 000 cells) and large (> 10 000 cells). Correlations in first column (over all species) are reproduced from Table 1 and 2 for convenient comparison.

	Range size				
	Over all species (n = 286)	Large ranges (n = 40)	Medium ranges (n = 117)	Small ranges (n = 132)	Partial, accounting for range size
Correlations of medians					
Number of cells	0.47	0.23	0.21	0.43	–
Longitude (degrees)	0.99	0.96	0.99	0.99	0.99
Latitude (degrees)	0.98	0.98	0.99	0.99	0.98
Elevation (m)	0.95	0.93	0.96	0.94	0.95
Mean annual temperature (C)	0.97	0.98	0.98	0.96	0.97
Annual precipitation (mm)	0.96	0.95	0.96	0.94	0.95
Diurnal temperature range (C)	0.94	0.87	0.94	0.95	0.94
Precipitation seasonality (no units)	0.94	0.92	0.93	0.97	0.95
Correlations of interquartile ranges					
Longitude (degrees)	0.89	0.89	0.66	0.73	0.75
Latitude (degrees)	0.84	0.68	0.77	0.77	0.83
Elevation (m)	0.71	0.81	0.73	0.71	0.71
Mean annual temperature (C)	0.72	0.70	0.72	0.76	0.73
Annual precipitation (mm)	0.66	0.81	0.62	0.68	0.66
Diurnal temperature range (C)	0.75	0.56	0.74	0.72	0.75
Precipitation seasonality (no units)	0.73	0.66	0.67	0.69	0.73
Correlations of interdecile ranges					
Longitude (degrees)	0.92	0.89	0.90	0.82	0.82
Latitude (degrees)	0.85	0.72	0.82	0.85	0.80
Elevation (m)	0.82	0.80	0.83	0.84	0.82
Mean annual temperature (C)	0.82	0.77	0.82	0.83	0.82
Annual precipitation (mm)	0.80	0.79	0.77	0.83	0.80
Diurnal temperature range (C)	0.84	0.52	0.83	0.81	0.83
Precipitation seasonality (no units)	0.85	0.68	0.83	0.83	0.85

(mainly Russia) was particularly depauperate in records from other sources (e.g. literature, atlases and website trip reports). La Sorte and Sumville (2020) report a similar scarcity of eBird surveys in northern Asia. The Map of Life (Jetz et al. 2012) indicates that coverage of birds (a measure of how well GBIF-mediated data characterize the makeup of avian assemblages averaged over all 150-km grid cells in a defined area) is only 0.4% for Russia, compared to 31% for the USA and over 60% for several Scandinavian countries (<<https://mol.org/indicators/coverage/>>; accessed 10 December 2018).

Latitude also differed noticeably for several species, although there was no apparent geographical pattern as with longitude; species that differed by more than 10° were scattered across the globe. The largest difference in latitude was also associated with the longest distance between range and point centroids, ~5400 km for Asian blue quail *Excalfactoria chinensis* (Fig. 6). This difference was due to the higher concentration of occupied points along the eastern Australia coast compared to the rest of the range, which reported extends northward into western India. As with longitude, this difference is most likely attributable to differences in data coverage, which vary from 17.5% in Australia (and up to 50% in its Pacific coastal grids) down to < 5% in India, southeastern Asia, Indonesia and the Philippines (<<https://mol.org/indicators/coverage/>>; accessed 10 December 2018).

Although over all species the correlation between range-based and point-based annual temperatures was similar to those of latitude and longitude, and nearly 1, the slope of the regression of point medians on range medians was shallower and the intercept higher (Table 1), such that points yielded slightly warmer values in colder regions (Fig. 3A). Again, we attribute this to the scarcity of points in northern and eastern Asia (e.g. Siberia) for a number of species compared to coverage of range maps. However, the largest individual species' difference between point- and range-based estimates of annual temperature reflects the coupling of temperature with elevation. Although only 125 km separates range versus point centroids of Sclater's monal *Lophoporus sclateri* in the mountain forests of the Himalayas (a region with a steep elevational gradient), the nearly 1300-m difference in median elevation (3222 m for ranges versus 1954 m for points) generates a 10°C difference in median temperatures. Overall, distance between centroids accounted for 11.4% of the variation in the magnitude (absolute value) of the difference between the two estimates ($r = 0.34$, $n = 286$).

Two of the species that displayed substantial differences (> 1000 mm) in annual precipitation estimates (Fig. 3B) both occurred on Borneo, an island with a very large precipitation gradient (1600–4200 mm). Despite having distribution centroids differing within the island by only 410 km,

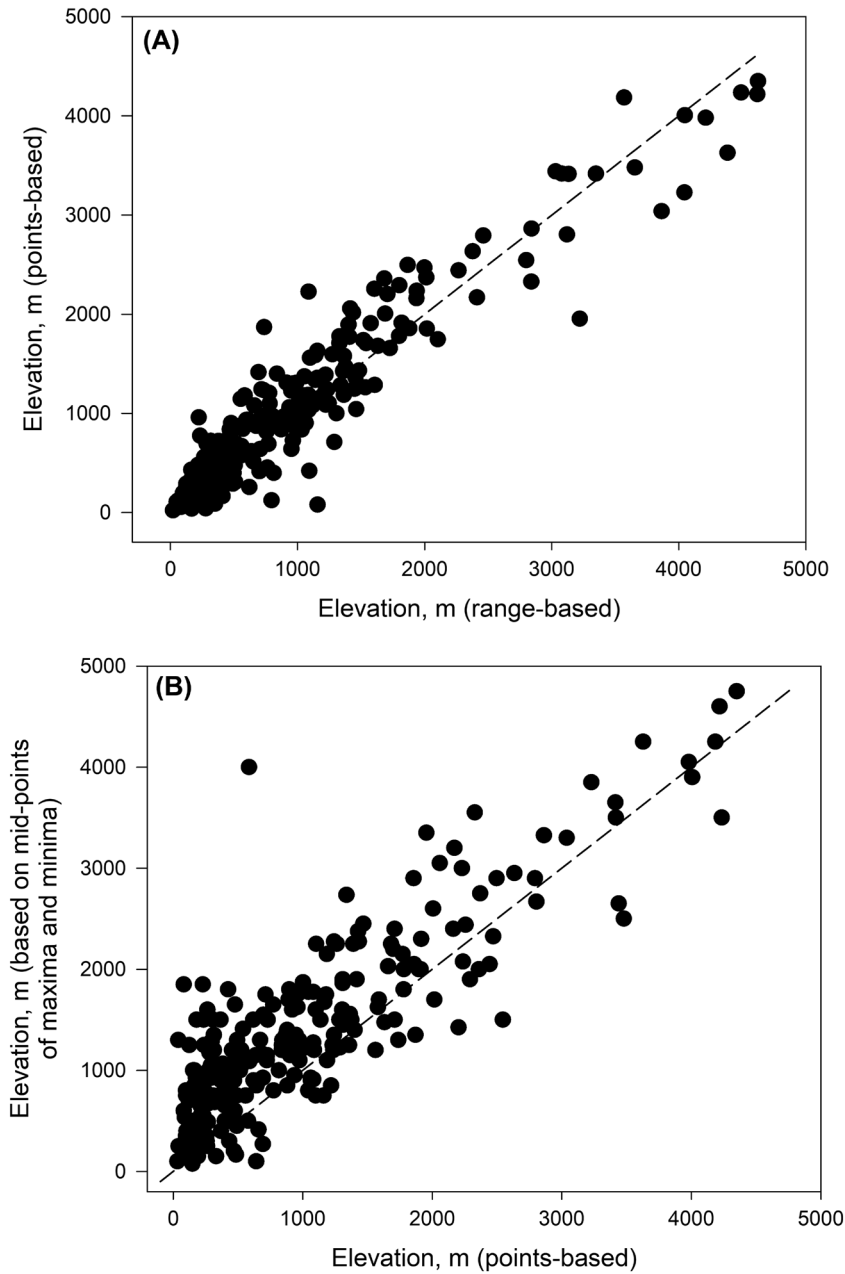


Figure 4. (A) Median elevation derived from points versus median derived from range maps. Dashed line represents 1:1 relationship. See Table 1 for correlation and regression coefficients. (B) Mid-point elevation derived from minimum and maximum (Quintero and Jetz 2018) versus median elevation derived from occupied points. Dashed line represents 1:1 relationship.

precipitation estimates for Bornean partridge *Arborophila hyperythra* differed by 1200 mm; crimson-headed partridge *Haematortyx sanguiniceps* centroids were separated by only 208 km, but precipitation differed by 1100 mm. Although elevation certainly plays a role in creating the precipitation gradient, these two species differ in median elevations based on ranges versus points of less than 200 m. Micronesian scrub-fowl *Megapodius laperouse* precipitation medians differed by 1200 mm, but these medians were based on only two occurrence cells versus five range cells, with occurrence cells from one island, range cells from another, 1300 km apart. This

particular large distance notwithstanding, distance between centroids overall accounted for only 1.6% of the difference in magnitude between estimates based on ranges versus points ($r = 0.13$, $n = 286$).

Without going into similar species-specific detail for other climate variables, in general those species that showed the largest deviations between ranges and occurrence points with respect to diurnal temperature range and precipitation seasonality (Fig. 3C–D) also tended to occur in areas with relatively steep spatial gradients in those environmental variables. Overall, the variation in the absolute difference in medians

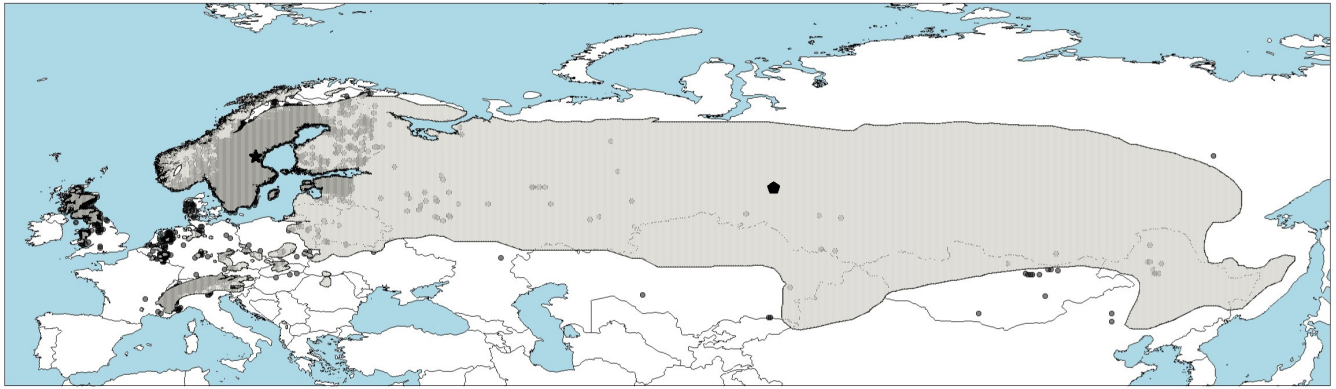


Figure 5. Range map of black grouse *Lyrurus tetrix* (from Birdlife; IUCN 2016) overlaying point localities (small gray or black dots, from GBIF.org 2018). Large star to west represents median latitude and longitude based on 3737 points (10' × 10' resolution); large pentagon to east represents median latitude and longitude based on rasterized range map (75 077 points). Distance between star and pentagon is ~3400 km. Image prepared in ModestR (Garcia-Rosello et al. 2013, 2014).

that was associated with the distance between centroids was 18.0% ($r=0.42$) for precipitation seasonality and 11.5% ($r=0.34$) for diurnal temperature range (both $n=286$).

Four species had differences between point-based and range-based elevations > 1000 m (Fig. 4A). Of these, three had point-based elevations characterized by only three points, whereas the fourth, with 95 points, manifest a distance of almost 3000 km between range versus point centroids. Thus, we attribute much of the discrepancy to inadequate sampling. But perhaps more importantly, differences between point-based and range-based estimates of median elevation were only about half the differences observed between point medians and minimum–maximum elevation mid-points, with the latter consistently higher. As implied by Quintero and Jetz (2018), species' detections might be biased downward as it may be more difficult to access higher elevations, leading to fewer observations there. On the other hand, mid-points may be biased upward to the extent observers may be more likely to report particularly high elevation observations of a species (i.e. beyond its customary range) as they may be deemed more noteworthy, and thus extend the maximum upward. However, regardless of any reporting biases, the difference between median and mid-point values for elevation is also influenced by geometry. Visualizing a mountain as a simple, tapering cone, it is apparent that it has a larger surface area (and hence more map cells with a value for elevation) below the mid-point between its base and its apex than above that mid-point. As a result, for any such tapering surface the median value of all cells will be less than the value that is the mid-point. Certainly, the magnitude of any difference between the two values (median versus mid-point elevation) for a species will depend on the topographic complexity of the landscape over which it occurs. Nevertheless, our observations are broadly consistent with this geometry.

The extreme outlier with respect to differences in elevation based on minimum–maximum mid-point versus median of occupied cells (Fig. 4B) was the rock ptarmigan *Lagopus muta*, a species described as occupying 'rocky tundra with

fairly sparse vegetation, or alpine summits' (HBW). Birdlife reports its elevational range as 2000–5000 m (3500 m mid-point), and Q&J gives values of 2000–6000 m (4000 m mid-point). In contrast, the median elevation of an occupied 10'-cell was 584 m (range 1–2807 m, $n=989$), and that of all cells within its range was 345 m (range 1–5455 m, $n=120\,499$). Conceivably, this could represent systematic misidentification arising from confusing this species with the similar-appearing sister taxon, the willow grouse, whose habitat is described as 'primarily Arctic tundra' (HBW). The two species overlap broadly, particularly in the northern parts of their ranges. For the willow grouse, Q&J provide a midpoint elevation of 490 m, compared to a median of 267 m estimated from point occurrence data (range 1–3646 m, $n=163\,171$). Nonetheless, regardless of the cause it is difficult to reconcile a discrepancy of 3400 m.

It is not surprising that there were exceptions to the general concordance of values in our analyses, given that we expect that not all points within a species' range will be occupied (in this case, the median number of cells occupied based on points was ~4% of the number based on ranges), and that points are more likely not to be occupied the more environmental conditions at the point differ from (in the sense of being less suitable than) others in the range. Given that the geographic locations (occupied cells) derived from the two different methods overlap considerably, at least some departures of point-based environmental variable values from range-based ones likely represent lack of occupancy of cells within ranges that are simply less suitable with respect to the variable. That the differences do not seem to be strongly influenced by geographical offset is further consistent with the notion that species' may have environment-restricted distributions within ranges, leading to 'holes' associated with less-suitable environment. Consistent with the notion that range maps might include inappropriate as well as appropriate environmental conditions, we observed variances (both inter-quartile and interdecile ranges) that were greater for range maps than occupied points for annual precipitation, diurnal

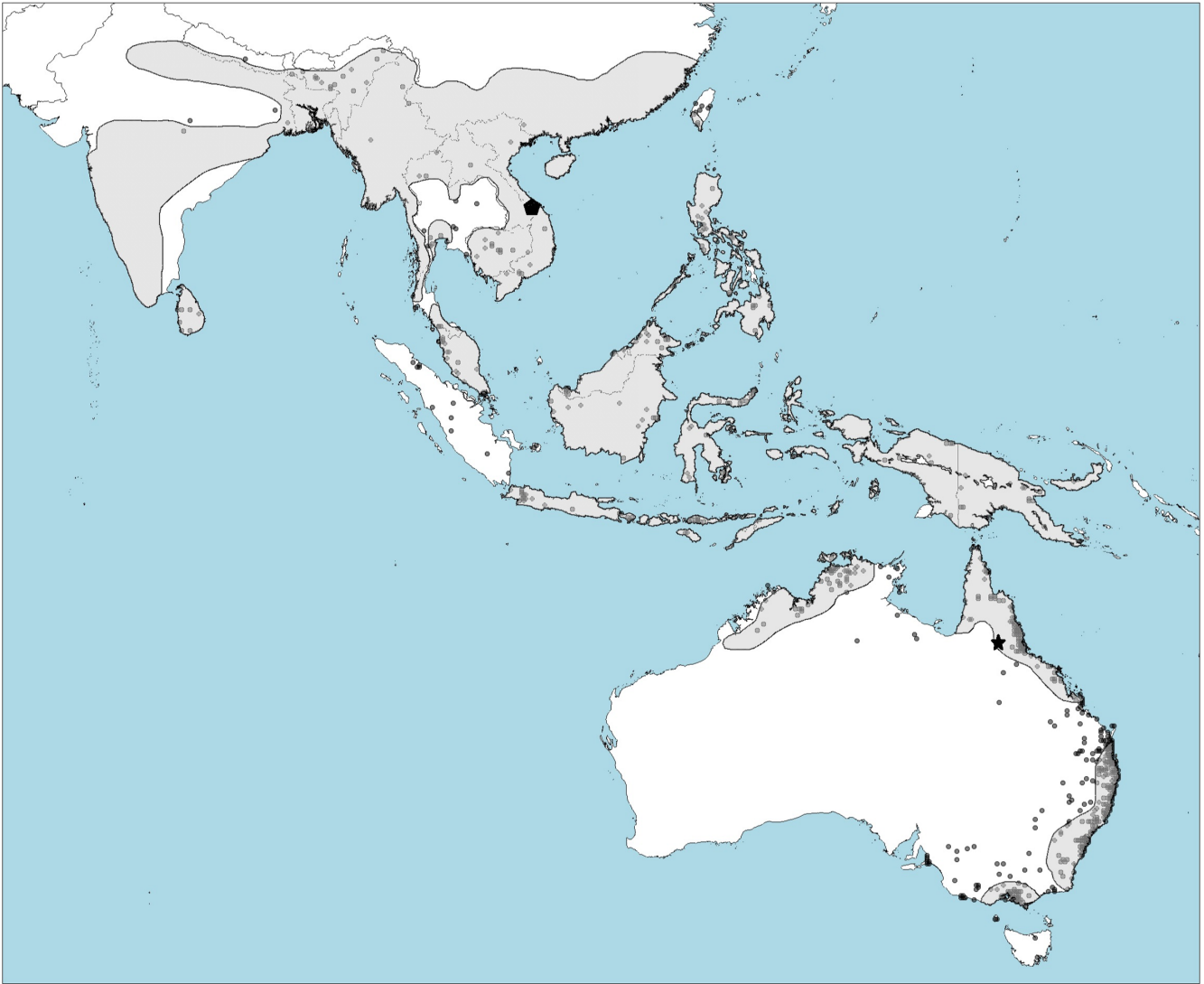


Figure 6. Range map of Asian blue quail *Excalfactoria chinensis* (from Birdlife; IUCN 2016) overlaying point localities (from GBIF.org 2018). Large star to southeast represents median latitude and longitude based on 638 points ($10' \times 10'$ resolution); large pentagon to north-west represents median latitude and longitude based on rasterized range map (24 233 points). Distance between star and pentagon is ~5400 km. Image prepared in ModestR (Garcia-Rosello et al. 2013, 2014).

temperature range and precipitation seasonality. However, points-based variances were generally greater than range-based ones for elevation and annual temperature, which we predicted could occur if we included occupied points that lay outside range map boundaries. To the extent that any observed difference between medians or variances seemed biologically relevant or important to a particular question at hand, it could be tested by taking repeated samples of N randomly selected cells within a species' range (N = the number of points for the species) to generate an expected median and variance value and their statistical distributions, to which the observed values could be compared (Fourcade 2016).

As we noted, not all deviations between range-based and point-based medians could be ascribed to a species' habitat specificity. In several instances it appeared that there was

simply an insufficient number of points to characterize adequately a species' value with respect to an environmental variable. Repeated random sub-samples of an environmental variable at a species' points across a range of sample sizes could provide some idea of the minimum N necessary to achieve a stable estimate of the median value. However, as this number is likely to vary as a function of the underlying geographical heterogeneity of the variable within the species' range at the scale sampled, sampling different environmental variables could produce different estimates of an adequate sample size. And if a species simply has too few points to begin with, subsampling is not feasible.

Although range-based and point-based estimates of median elevation of species were highly correlated, these estimates differed substantially from those based on mid-points

of the reported span (minimum–maximum) of elevations occupied. Although it is certainly possible that the discrepancy is due to inadequate sampling (i.e. fewer points) at higher elevations within a species' range of occurrence, an alternative, that the distribution of a species' occurrences along an elevational gradient is not symmetrical (and hence the mid-point is not a suitable measure of central tendency, as we argue above based on topographic geometric considerations), seems plausible as well. For example, a lowland species may have a reported range of 0–1000 m, where it may have only been seldom observed at 1000 m but is widely distributed at sea level. Compared to the mid-point elevation of 500 m, this is an immediate skew of nearly 500 m. Clearly the frequency distributions of species' points in our sample were generally skewed, with relatively more at lower versus higher elevations. In this event, the median will be a more suitable measure of the central tendency of the distribution than will the mid-point, which better characterizes the central tendency of a symmetrical distribution. Nevertheless, the best resolution of this discrepancy is also likely the most difficult, a more thorough sampling and reporting of species' distributions along elevational gradients.

To what extent does this elevational discrepancy even matter? To the extent that one uses elevation as a proxy for, say, temperature, if one derives the actual values of environmental variables (e.g. climate) associated with a species distribution from either range maps or occurrence points then associating elevation (meters) with a species' distribution may be irrelevant. There is little point in determining the value of a proxy if one has in hand the values of the variables for which the proxy is a surrogate. Moreover, although temperature globally declines with elevation, it does so with an interaction with latitude; average annual temperature at 1500 m at 20°N is different from that at the same elevation at 50°N. However, other physical environmental features that may influence life history traits in animals and plants, such as air pressure and oxygen availability, are not latitude-dependent, and having a more accurate estimate of a species' elevational distribution becomes more important (Körner 2007).

Concluding remarks

For most species, we conclude that using points of known occurrence will likely provide more reliable environmental data than overlaying range maps, although differences between the two may be slight. Perhaps most importantly, use of points can obviate the 'porosity' issue of range maps, and its creation of false-positives of occurrences. The performance of several commonly used species distribution modeling techniques that employ presence/absence data have been shown to be more adversely affected by false-positives than by false-negatives (Fernandes et al. 2019), and this may carry over to other types of models as well. Another significant advantage to points is deflating the overestimation of a species' elevational distribution based on mid-points between minimum and maximum extents. Likewise, by retaining points that occurred outside of range map boundaries one

includes peripheral or disjunct populations that range maps often omit. We acknowledge, however, that the 500-km limit we applied to retaining out-of-range points was arbitrary and that different limits would possibly produce different results, although this effect was small in the analyses of Alhajeri and Fourcade (2019). We further note that even with relatively dense sampling of a species' distribution, objectively determining the exact boundaries of a species' range can be a fraught process (Fortin et al. 2005). Indeed, the observation that some variables had higher range-based variances whereas others had higher points-based variances suggests potentially poor accuracy in determining the exact limits of species' distributions (Alhajeri and Fourcade 2019). Rasterizing species' occurrences to simply presence in a 10' map cell regardless of the number of occurrences reduces the 'hot spot' effect by eliminating redundant observations, although it does not necessarily remove other forms of non-randomness in the distribution of points (e.g. along roads). Moreover, the size of our map cells absorbs some imprecision in geographical coordinates. What using points does not do, however, is increase the amount of data available for species with few observations (which often have small ranges as well), nor does it rectify gross inconsistencies in geographical reportage of observations of a species. Nevertheless, macroecological analyses based on species' occurrences will continue to improve as the challenges of increasing the coverage and reducing the biases of observational data are met (Callaghan et al. 2019).

Data availability statement

Species medians and interquartile and interdecile ranges for all variables extracted from both point occurrences and range maps will be available from the Dryad Digital Repository: <<https://doi.org/10.6086/D1Z396>> (Rotenberry and Balasubramaniam 2020).

Acknowledgements – We express gratitude to the thousands of contributors to the Global Biodiversity Information Facility, including eBird, and the International Union for the Conservation of Nature for the data that make macrogeographical analyses such as our possible. We thank A. T. Peterson for comments that helped clarify our presentation. We also thank Seshadri Rajagopal for assistance with GIS issues. The Univ. of Minnesota Dept of Ecology, Evolution and Behavior provided administrative support for this project.

Conflicts of interest – The authors declare no conflicts of interest.

Author contributions – JTR and PB conceived the project; JTR collected and analyzed the data; JTR prepared the initial draft and JTR and PB revised the manuscript.

References

- Alhajeri, B. H. and Fourcade, Y. 2019. High correlation between species-level environmental data estimates extracted from IUCN expert range maps and from GBIF occurrence data. – *J. Biogeogr.* 46: 1329–1341.

- Balasubramaniam, P. and Rotenberry, J. T. 2016. Elevation and latitude interact to drive life-history variation in precocial birds: a comparative analysis using galliformes. – *J. Anim. Ecol.* 85: 1528–1539.
- Beck, J. et al. 2014. Spatial bias in the GBIF database and its effect on modelling species' geographic distributions. – *Ecol. Inform.* 19: 10–15.
- BirdLife International and NatureServe 2015. Bird species distribution maps of the world. Ver. 2.0. – BirdLife International, Cambridge, UK and NatureServe, Arlington, USA, <<http://datazone.birdlife.org/species/requestdis>> accessed January 2015.
- Boakes, E. H. et al. 2010. Distorted views of biodiversity: spatial and temporal bias in species occurrence data. – *PLoS Biol.* 8: e1000385.
- Boyce, A. J. et al. 2015. Clutch size declines with elevation in tropical birds. – *Auk* 132: 424–432.
- Boyle, W. A. et al. 2016. Patterns and drivers of intraspecific variation in avian life history along elevational gradients: a meta-analysis. – *Biol. Rev.* 91: 469–482.
- Callaghan, C. T. et al. 2019. Improving big citizen science data: moving beyond haphazard sampling. – *PLoS Biol.* 17: e3000357.
- del Hoyo, J. et al. (eds) 2017. Handbook of the birds of the world alive. – Lynx Edicions, <www.hbw.com/> accessed March 2018.
- Elith, J. and Leathwick, J. 2009. Species distribution models: ecological explanation and prediction across space and time. – *Annu. Rev. Ecol. Evol. Syst.* 40: 677–697.
- Fernandes, R. F. et al. 2019. Effects of simulated observation errors on the performance of species distribution models. – *Divers. Distrib.* 25: 400–413.
- Fick, S. E. and Hijmans, R. J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. – *Int. J. Climatol.* 37: 4302–4315.
- Fortin, M.-J. et al. 2005. Species geographic ranges and distributional limits: pattern analysis and statistical issues. – *Oikos* 108: 7–17.
- Fourcade, Y. 2016. Comparing species distributions modelled from occurrence data and from expert-based range maps. Implication for predicting range shifts with climate change. – *Ecol. Inform.* 36: 8–14.
- Fourcade, Y. et al. 2014. Mapping species distributions with MAX-ENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. – *PLoS One* 9: e97122.
- Franklin, J. 2010. Mapping species distributions: spatial inference and prediction. – Cambridge Univ. Press.
- Franklin, J. et al. 2014. Linking spatially explicit species distribution and population models to plan for the persistence of species under global change. – *Environ. Conserv.* 41: 97–109.
- García-Roselló, E. et al. 2013. ModestR: a software tool for managing and analyzing species distribution map databases. – *Ecography* 36: 1202–1207.
- García-Roselló, E. et al. 2014. Using ModestR to download, import and clean species distribution records. – *Methods Ecol. Evol.* 5: 703–713.
- Gardner, A. S. et al. 2019. Climatic predictors of species distributions neglect biophysiological meaningful variables. – *Divers. Distrib.* 25: 1318–1333.
- GBIForg 2018. GBIF home page. – <www.gbif.org>.
- Gill, F. and Donsker, D. (eds) 2017. IOC world bird list (V7.3). – <www.worldbirdnames.org/ioc-lists/crossref/>.
- Global Digital Elevation Model 2004. U.S. Geological Survey, EROS Data Center Distributed Active Archive Center (EDC DAAC). – <https://developers.google.com/earth-engine/datasets/catalog/USGS_GTOPO30>.
- Graham, C. H. and Hijmans, R. J. 2006. A comparison of methods for mapping species ranges and species richness. – *Global Ecol. Biogeogr.* 15: 578–587.
- Graham, C. H. et al. 2004. Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. – *Evolution* 58: 1781–1793.
- Guisan, A. et al. 2013. Predicting species distributions for conservation decisions. – *Ecol. Lett.* 16: 1424–1435.
- Hijmans, R. J. 2015. raster: geographic data analysis and modeling. – R package ver. 2.5-2, <<http://CRAN.R-project.org/package=raster>>.
- Hurlbert, A. H. and White, E. P. 2005. Disparity between range map- and survey-based analyses of species richness: patterns, processes and implications. – *Ecol. Lett.* 8: 319–327.
- Hurlbert, A. H. and Jetz, W. 2007. Species richness, hotspots and the scale dependence of range maps in ecology and conservation. – *Proc. Natl Acad. Sci. USA* 104: 13384–13389.
- IUCN 2016. The IUCN Red List of threatened species. – <www.iucnredlist.org/resources/spatial-data-download>.
- Jacobs, C. and Zipf, A. 2017. Completeness of citizen science biodiversity data from a volunteered geographic information perspective. – *Geo-spatial Inform. Sci.* 20: 3–13.
- Jenkins, C. N. et al. 2013. Global patterns of terrestrial vertebrate diversity and conservation. – *Proc. Natl Acad. Sci. USA* 110: E2602–E2610.
- Jetz, W. et al. 2008. The worldwide variation in avian clutch size across species and space. – *PLoS Biol.* 6: 2650–2657.
- Jetz, W. et al. 2012. Integrating biodiversity distribution knowledge: toward a global map of life. – *Trends Ecol. Evol.* 27: 151–159.
- Körner, C. 2007. The use of 'altitude' in ecological research. – *Trends Ecol. Evol.* 22: 569–574.
- La Sorte, F. A. and Somveille, M. 2020. Survey completeness of a global citizen-science database of bird occurrence. – *Ecography* 43: 34–43.
- Lack, D. 1947. The significance of clutch-size, parts 1 and 2. – *Ibis* 89: 302–352.
- McPherson, J. M. and Jetz, W. 2007. Type and spatial structure of distribution data and the perceived determinants of geographical gradients in ecology: the species richness of African birds. – *Global Ecol. Biogeogr.* 16: 657–667.
- Meyer, C. et al. 2016. Multidimensional biases, gaps and uncertainties in global plant occurrence information. – *Ecol. Lett.* 19: 992–1006.
- Peterson, A. T. et al. 2016. Assumption-versus data-based approaches to summarizing species' ranges. – *Conserv. Biol.* 32: 568–575.
- Qian, H. T. et al. 2018. Incomplete species lists derived from global and regional specimen-record databases affect macroecological analyses: a case study on the vascular plants of China. – *J. Biogeogr.* 45: 2718–2729.
- Quintero, I. and Jetz, W. 2018. Global elevational diversity and diversification of birds. – *Nature* 555: 246–250.
- Roff, D. A. 1992. The evolution of life Histories: theory and analysis. – Chapman and Hall.
- Rotenberry, J. T. and Balasubramaniam, P. 2020. Data from: Connecting species' geographical distributions to environmental variables: range maps versus observed points of occurrence. –

- Dryad Digital Repository, <<https://doi.org/10.6086/D1Z396>>.
- SAS Inst. 2013. SAS Ver. 9.3. – SAS Inst., Cary, NC.
- Scott, J. M. et al. 1993. Gap analysis: a geographic approach to protection of biological diversity. – *Wildl. Monogr.* 123: 1–41.
- Sullivan, B. L. et al. 2009. eBird: a citizen-based bird observation network in the biological sciences. – *Biol. Conserv.* 142: 2282–2292.
- Sullivan, B. L. et al. 2014. The eBird enterprise: an integrated approach to development and application of citizen science. – *Biol. Conserv.* 169: 31–40.
- van Brummelen, G. R. 2013. Heavenly mathematics: the forgotten art of spherical trigonometry. – Princeton Univ. Press.
- Yesson, C. et al. 2007. How global is the global biodiversity information facility? – *PLoS One* 2: e1124.
- Zar, J. H. 1984. Biostatistical analysis, 2nd ed. – Prentice-Hall.
- Supplementary material (available online as Appendix ecog-04871 at <www.ecography.org/appendix/ecog-04871>). Appendix 1.